Spatial Hearing, Attention and Informational Masking

Gerald Kidd
TRUSTEES OF BOSTON UNIVERSITY

10/26/2015
Final Report

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
10/11/2015	Final	May 2012 - April 2015

4. TITLE AND SUBTITLE
Spatial Hearing, Attention and Informational Masking in Speech Identification

5a. CONTRACT NUMBER
N/A

5b. GRANT NUMBER
FA9550012-1-0171

5c. PROGRAM ELEMENT NUMBER
N/A

6. AUTHOR(S)
Gerald Kidd

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Trustees of Boston University
882 Commonwealth Avenue
Boston, MA 02215

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USAF, AFRL
875 North Randolph Street, Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION A

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The research completed during the past award period concerned the development and evaluation of a new means for representing the speech stimulus, called "pointillistic speech," and the role of listener expectation in speech understanding in adverse acoustic environments. The topics addressed using pointillistic processing included informational masking and the covert embedding of secondary messages in pointillistic speech and other sounds. The primary accomplishment was obtaining a better understanding of speech communication under masked conditions while a secondary accomplishment was the testing and refinement of covert messaging using pointillistic speech. In the latter case, possible applications for increasing information transfer of audio signals and acoustic steganography were explored. In the area of predictability and expectation, the reslience of streams of speech to perceptual and cognitive intrusions by competing sounds was explored on a linguistic level by varying syntactic structure and by using more formal means for varying predictability by constructing target sequences of sounds from transistion matrices (i.e., Markov chains). This approach was successful in improving our understanding of the processing of streams of speech masked by other speech.

15. SUBJECT TERMS
Speech intelligibility; informational masking; spatial hearing; acoustic steganography; pointillistic speech

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Gerald Kidd, Jr., Ph.D.
				12	19b. TELEPHONE NUMBER *(Include area code)* 617-353-7360

Reset

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18
Adobe Professional 7.0

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

FINAL REPORT: "Spatial hearing, attention and informational masking in speech identification"

The following report describes our work during the award period from May 1, 2012, through April 30, 2015. The progress report is organized according to the objectives stated in the proposal and includes short narrative summaries with representative figures and a list of publications and presentations supported by this award. A more complete description of the findings presented here in abbreviated form can be found in the referenced publications (pdfs available upon request) or in more extensive written summaries of unpublished work (also available upon request) including the summaries given at the Sensory Information Systems program reviews in October, 2013, and 2014.

I. Collaboration

Significant portions of the work supported by this award have been conducted through the collaborative efforts of the research groups at Boston University and WPAFB. These two groups and associated collaborators routinely hold joint discussions of research at the annual Acoustical Society of America Spring Meeting, at the Midwinter Research Meeting of the Association for Research in Otolaryngology, and in Boston at the annual Binaural Bash conference sponsored by the Hearing Research Center at Boston University. The specific projects upon which we have collaborated are evident in the co-authored articles and presentations at scientific conferences listed in a following section. Both of the broad areas of work proposed here: using highly-quantized representations of speech signals as a means for examining informational masking and the role of predictability and a priori knowledge in speech stream perception, are areas that have been/will be addressed by collaborative work between these two groups and our associated consultants and other contributors. These projects reflect shared interest in the research questions addressed by this proposal and in their relevance to the scientific mission of AFOSR. Our plan going forward is to continue these collaborative efforts very much in the same manner as in previous years.

II. **Progress Achieving the Objectives of the Research Plan**

The work proposed in this application continues and significantly extends the work undertaken during the previous award period. Some of the methods we propose to use were developed during the past award period and some new applications emerged that guide our specific interests. Most importantly, though, the work that was completed from 2012-2015 achieved its intended broad goal: our understanding of the auditory and cognitive processes underlying human communication in multiple-talker sound fields (i.e., the "cocktail party problem") has increased dramatically and leads directly to the work that is planned and described in the current proposal. The following section reviews our progress toward the objectives described in the last application.

A. To obtain a better understanding of the mechanisms of auditory masking using "pointillistic" processing of speech

The advantage of pointillistic representations of speech is that it allows for precise control of the information that is available to the listener while also (potentially) preserving the essential aspects of speech for intelligibility. This high degree of control can be used to determine how well specific speech features are maintained under masked conditions and - on an even more basic level - what acoustic information is needed to form the minimal representations of these features. Minimal/sparse representations could be used to reduce the amount of data required to code speech (potentially improving transmission efficiency for specific applications) or for enhancing speech in ways that make it less susceptible to masking (e.g., emphasizing the important features during transmission).

The first two projects discussed in this section provide examples and summaries of our work in this area. The first project, in collaboration with Professor Lori Holt of Carnegie-Mellon University (consultant on grant during previous award period), examined the idea of sparse representations of one specific phonemic contrast that forms the basis for the categorical perception of the consonant-vowel (CV) sequence /ba/-/da/-/ga/. The approach was to take an established natural sequence of CV pairs that formed an acoustic continuum that, in this case, was distinguished by the formant transitions (principally the second formant) - the principal feature upon which the /ba/-/da/-/ga/ categorical judgment is based.
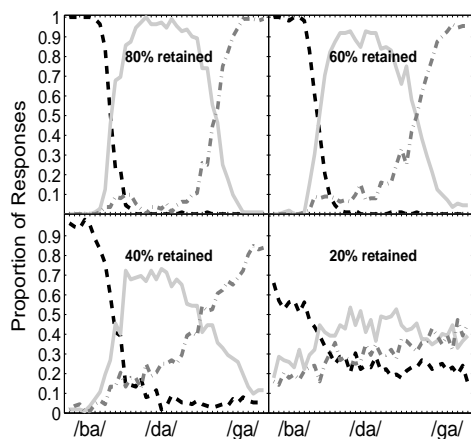
Figure 1 (left) shows human performance (proportion of categories chosen) in labelling consonant-vowel pairs as /ba/, /da/ or /ga/ in a 3-alternative forced-choice procedure (without feedback) as a function of the acoustic variable signaling phoneme identity. The primary physical variable was the slope of the second formant transition and the details of the methods used to construct this stimulus continuum are given in Stephens and Holt (2011). The categorical nature of the pattern of results is seen in the sharp change from nearly 100% identification of one phoneme (transitioning to near 0% identification of that phoneme) to nearly 100% identification of the next phoneme along the continuum - despite the fact that each step along the abscissa is an equal change in the physical variable. The category boundaries for natural/unprocessed speech are not displayed because they are nearly identical to those shown for 80% retention of the points representing the phonemes (upper left panel of **Figure 1**). The purpose of this work was to determine how sparse a pointillistic representation could be and still accurately convey the feature underlying phoneme identity. This figure presents a portion of that work, and indicates that reasonably intact category boundaries are possible with only about 60% of the points remaining after applying a process that randomly removed points within the time-frequency matrix representing the stimulus. A parallel manipulation - removing the points below a particular amplitude that is specified by the experimenters - indicated that the category boundaries could be preserved if only the points falling within the upper 20% of the amplitude range were retained and the remainder discarded. These manipulations support the hypothesis that pointillistic speech can be used to convey speech features accurately with sparse representations. Note that these manipulations used random or arbitrary rules to remove points. We currently are exploring ways of minimizing the pointillistic representations so that the feature of interest is conveyed with the minimum possible number of points guided by our a priori knowledge of the speech feature or by a algorithm that uses the results of perceptual experiments that determine the weighting of points as they contribute to intelligibility (i.e., which points are most important for distinguishing /ba/ from /da/; those points are retained while others not important for determining the category are discarded). This work is ongoing.

A second project was completed that relied on pointillistic speech as the stimulus. This project was not originally planned under the objectives stated in the previous proposal, but rather grew out of a "discovery" that could have practical applications of interest to AFOSR. This work was first reported at the annual review meeting of the Sensory Information Systems program of AFOSR at Fort Walton Beach, FL, in October, 2014 and is described in Kidd and Mason (2015). The basic idea from an acoustic analysis standpoint is shown in the two panels of **Figure 2** (below).
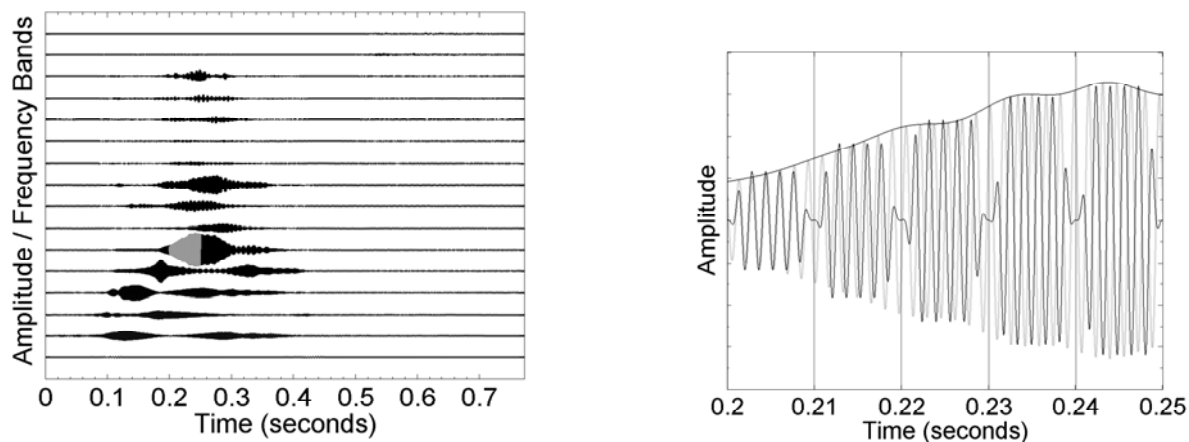
The left panel shows the initial stage of the spectro-temporal analysis that leads to pointillistic representations. The speech signal is filtered into 16 contiguous frequency bands. The Hilbert envelope and phase (instantaneous frequency) functions are obtained for each band to provide the values for the points that substitute for the original waveform. Within each band and time segment, the averages of the envelope magnitude and instantaneous frequency are computed and are used to generate a pure-tone "point" that replaces the corresponding segment of the original waveform. This process is illustrated in the right panel of **Figure 2**. The lightly shaded section of frequency band six (low to high) in the left panel is expanded and replotted in the right panel. Here the envelope (smoothed black curve along the top of the waveform) and waveform fine structure (light gray) of the original signal and the five 10-ms duration ramped pure-tone points (dark waveforms) that replace it are depicted. This analysis/synthesis is performed on all time-frequency units of the speech stimulus yielding a complete pointillistic representation that past work has shown to be highly intelligible (Kidd et al., 2009) although it lacks strong intonation and some other features of natural-sounding speech. The key aspect of this process for the current discussion is the (lack of) importance of the starting phases of the individual points. Speech intelligibility experiments performed in our laboratory during the previous award period have demonstrated that variations in the starting phases of the points - unlike scrambling the phase of natural speech - do not affect intelligibility. What this means is that the phases then become a "free parameter" that may be arbitrarily - or deliberately - specified without degrading the intelligibility of the speech signal (which will be called "the *primary message*"). Subsequent work on this project has found that a *secondary message* may be coded by the phases of the individual points; for example a starting phase of 0 radians is assigned a binary value of 0 and a starting phase of π radians is assigned a binary 1. If those values may be recovered by signal processing then the phases may be converted to bits forming binary words (e.g., a character specified by ASCII code or a word in a table look up in a predetermined lexicon). So in each time segment in each band one bit may be coded by the phase of the pure-tone point without altering the intelligibility of the primary message. A secondary message may then be covertly coded in the bit pattern of the speech signal. We have pursued this line of work with a goal of determining various types of natural-pointillistic hybrid speech that retain high intelligibility, good sound quality, and (relatively) high rate of transmission of the secondary message (in the range of hundreds of words/sec).

Three examples are shown in **Figure 3** (right) for different types of hybrid speech. For all three examples, near 100% correct intelligibility and near 100% correct recovery of the secondary message (phase-coded ASCII characters) were achieved. Each example has eight higher-frequency pointillistic bands added to natural speech. The conditions shown are (top to bottom): natural speech low-passed at 4 kHz, pointillistic bands from roughly 1.4-8 kHz; broadband natural speech, pointillistic bands from 5-12 kHz; and broadband natural speech, pointillistic bands from 8-20 kHz. Work on this project continues.

In the previous two projects, pointillistic speech was presented to the observer in isolation rather than in masked conditions. In the next project, we employed a closely related approach based on "ideal binary masking" or "ideal time-frequency analysis" (ITFA) that has been used to separate EM from IM in speech mixtures (Brungart et al., 2006). This approach is closely related to pointillistic processing in that it yields a highly-quantized analysis of speech that is reduced to minimal time-frequency (T-F) units. The ITFA approach yields speech representations that can be used as stimuli in speech recognition experiments - just as with pointillistic speech - that allow a high degree of control of the stimulus. Our initial work in this area that has provided us with the background and motivation for portions of the work in the current proposal and stems in part from the following joint project between the group at BU led by Dr. Virginia Best and Ms. Christine Mason, the research group at WPAFB led by Dr. Nandini Iyer, and the group at Walter Reed Medical Center led by Dr. Douglas Brungart. The goal of the project was to explore the EM-IM contrast for two different types of speech materials and tests that, according to our initial hypothesis, should differ in the degree to which spatial separation of sources provides a release from IM. Portions of this work are described in a manuscript in press (Best et al., 2015). Furthermore, the experimental approach sought to separate acoustic "better ear" cues available in moments where the amplitude of the masker speech was low while the target speech was correspondingly relatively high, producing brief "glimpses" of the target from perceptual segregation due to internally-generated binaural processing.
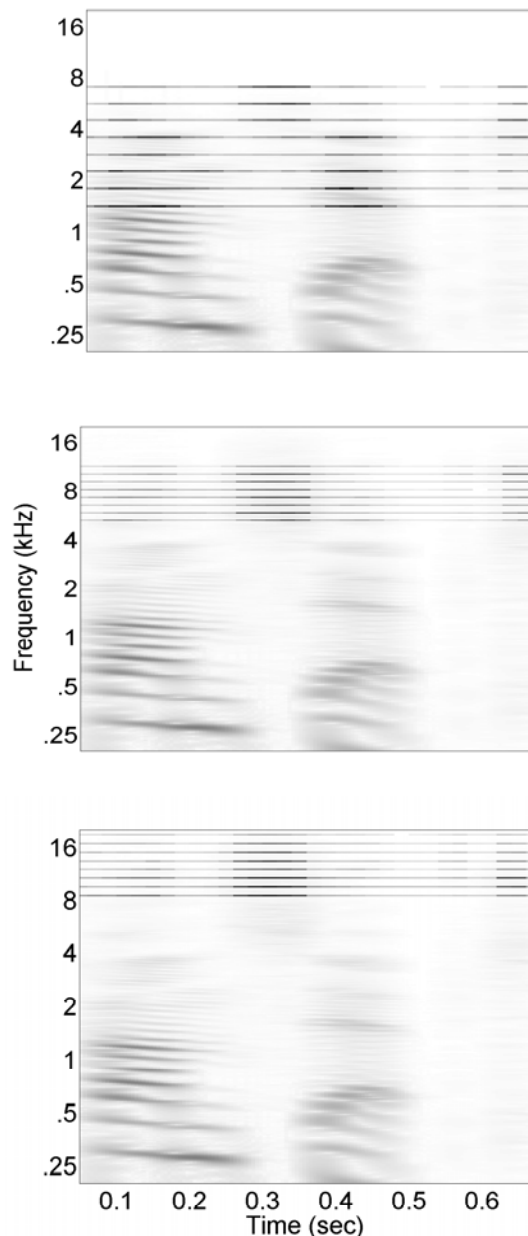
For each type of speech test (the "Modified Rhyme Test" - MRT - and the "Coordinate Response Measure" - CRM) the key comparison was between speech identification performance in a condition with no apparent spatial separation of sources or binaural cues (colocated at 0° azimuth) vs. two cases where binaural processing was applied. In one case, called "*binaural*," the two maskers were natural speech (unprocessed) and were presented spatialized (via head-related transfer functions, HRTFs) at ±60° azimuth. In the second case (called "*better-ear*"), the processing described in detail in Brungart and Iyer (2012) was applied to the stimuli: The target was always presented in the center at 0° azimuth. Left and right ear masker signals were bandpass filtered by a 128-channel gammatone filterbank (80–5000 Hz). The time-domain outputs from each channel were then divided into 20-ms segments (with 50% overlap) and multiplied by a 20-ms raised-cosine window. Because the HRTFs were symmetric across the median sagittal plane, the target signal was always identical in the two ears. Thus it was possible to identify the "better ear" as the ear with the lower total root-mean-square (rms) energy within each windowed T-F segment. For every time/frequency segment, the *better ear* was chosen from the binaural stimulus, the better-ear elements (one for each T-F unit) were combined, and the resynthesized mixture was presented diotically.

The results from this experiment are shown in **Figure 4** (right). For the MRT test and materials, which, according to our interpretation are relatively low in IM, *better-ear* and *binaural* conditions yielded about the same advantage re. colocated: a "benefit" of approximately 5 dB. For the CRM test and materials, which tend to produce large amounts of IM, the advantage for *better-ear* was about 6 dB, while the advantage for *binaural* presentation increased to about 9 dB.
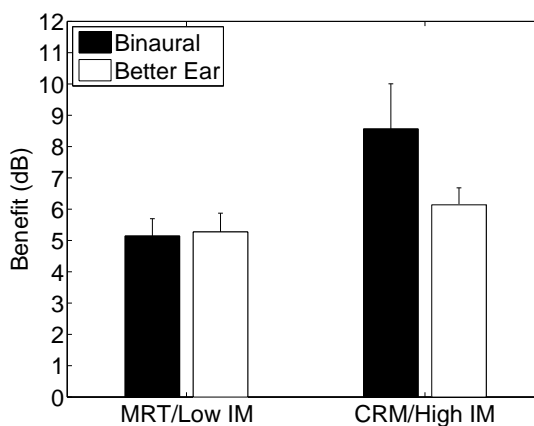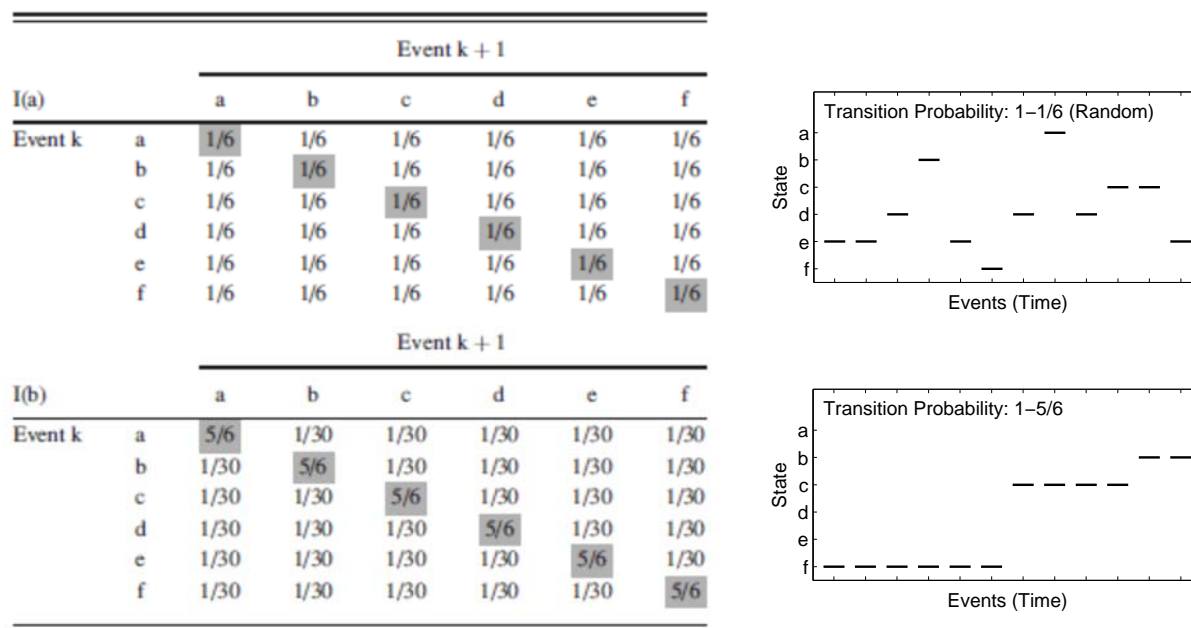

Thus, the conclusion was that relying on fine-grained T-F glimpses corresponding to moment-by-moment acoustic "better-ear" advantages can provide a significant benefit when the sound sources are symmetrically separated in azimuth. This acoustically-determined benefit is about the same for cases where EM dominates performance. However, when the listening situation is high in IM, the benefit from the apparent separation of sources that is achieved through binaural listening is significantly greater; in this case the difference was about 3 dB greater than for *better-ear* alone.

B. To examine the role of listener expectation in stream formation and maintenance

The premise upon which this objective is based is that human listeners exploit a priori knowledge about source and message probabilities to successfully perform sound source segregation, selection and speech recognition in multiple-source sound fields. To examine the merits of this idea, we completed two major projects that tested aspects of this putative process. In the first project, which was a collaborative effort among the group at BU led by Dr. Kidd, Dr. Gregory Wakefield of the University of Michigan (consultant on grant during previous award period), and Dr. Eric Thompson of Ball Aerospace & Technologies Corporation/WPAFB

(described in Kidd et al., 2013), the sensitivity of listeners to statistical dependencies in sequences of nonspeech sounds was systematically examined. The basic idea here is that, in order for listeners to exploit a priori information in performing stream segregation and maintenance, we must first determine that listeners are, indeed, able to detect the presence and judge the strength of statistical dependencies among the elements of sound sequences. To test this hypothesis empirically, sequences of nonspeech sounds were presented that were chosen from transition matrices to form Markov chains (e.g., Rabiner, 1998). The underlying perceptual dimensions along which the stimuli varied were pitch (pure-tone frequency) and spatial location (varied according to interaural time differences or "ITDs"). Examples of sound sequences drawn from six states from either of two signal variables (frequency/pitch and ITD/spatial location) and the transition matrices from which they were constructed are shown in **Figure 5** (below).

I(a)

		Event k + 1					
		a	b	c	d	e	f
Event k	a	1/6	1/6	1/6	1/6	1/6	1/6
	b	1/6	1/6	1/6	1/6	1/6	1/6
	c	1/6	1/6	1/6	1/6	1/6	1/6
	d	1/6	1/6	1/6	1/6	1/6	1/6
	e	1/6	1/6	1/6	1/6	1/6	1/6
	f	1/6	1/6	1/6	1/6	1/6	1/6

I(b)

|		Event k + 1					
		a	b	c	d	e	f
Event k	a	5/6	1/30	1/30	1/30	1/30	1/30
	b	1/30	5/6	1/30	1/30	1/30	1/30
	c	1/30	1/30	5/6	1/30	1/30	1/30
	d	1/30	1/30	1/30	5/6	1/30	1/30
	e	1/30	1/30	1/30	1/30	5/6	1/30
	f	1/30	1/30	1/30	1/30	1/30	5/6

The left panel (table) shows two example transition matrices comprising six states A-F which correspond to six values of pure-tone frequency or ITD. The right panel is an illustration of the sequences of states that are "draws" from the two matrices and that would be used to present the sounds in the psychophysical discrimination experiment. Here the stimulus drawn from the non-uniform transition matrix (lower right panel) has fewer transitions than the stimulus drawn from the uniform (random) transition matrix, so the correct observer response would be to choose the stimulus in the right-side lower panel.

The results from the psychophysical experiment testing the discriminability of sequential dependencies for the two perceptual variables - pitch and location - are shown in **Figure 6** (right). This figure shows percent correct discrimination for sequences of sounds (2-interval 2-alternative forced-choice) based on the difference in the strength of the sequential dependency generating the sequences (diagonal entries in matrix). Here we found that listeners were quite sensitive to the statistical properties of sequences of sounds when the "states" in the transition matrices were frequency or apparent location (consistently better performance for frequency) but performed notably less well than the Ideal Observer; that monitoring two concurrent streams separated by frequency and by ear of presentation was possible and resulted in performance well above chance (not shown) but was poorer than comparable performance when the streams were successively presented. Performance was not affected by the two durations tested. Overall, this work was interpreted as supporting the underlying hypothesis that statistical dependencies can be exploited by listeners in forming and maintaining perceptual streams. Furthermore, this work served as a successful feasibility test for an empirical approach (using Markov chains as stimuli) to assess perceptual stream formation and maintenance under masked conditions. An Ideal Observer approach to modeling also seems to be useful in helping to understand human limitations on performance.
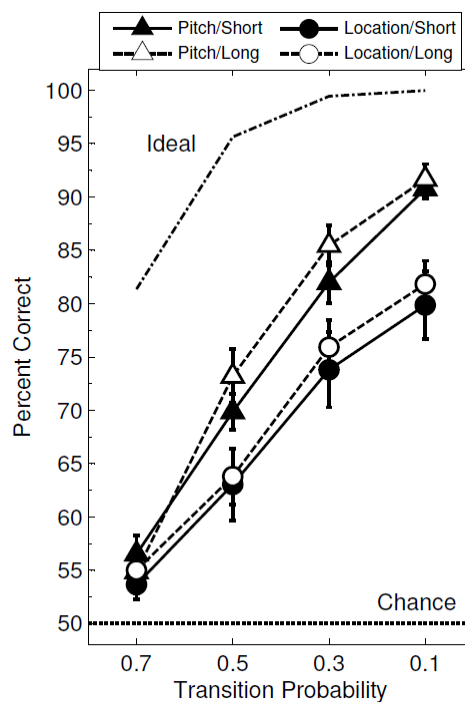
In the second major project under this objective (complete description in Kidd et al., 2014), the goal was to test the hypothesis that listener expectation could be beneficial in maintaining the focus of attention on a target stream of speech in competition with other masking streams of speech. This work tested the hypothesis - similar to that examined in the preceding work above - that a priori knowledge about the probability of occurrence of the elements comprising perceptual streams can be exploited by the observer to segregate and selectively attend to one specific source among competing sources. However, in this project the target sequence was intelligible speech as were the competing masking sources. The a priori information that was the primary controlled variable was the conformance to a known and valid syntactic structure: in the critical comparison, the target speech was syntactically correct vs. syntactically incorrect (random words). Two other main variables were also tested: the predictability of low-level segregation cues (reliable spatial *location* of target source [open symbols] or target talker *voice* [filled symbols]) and the interaction between syntactic structure and IM value of the masker (speech vs. noise control).
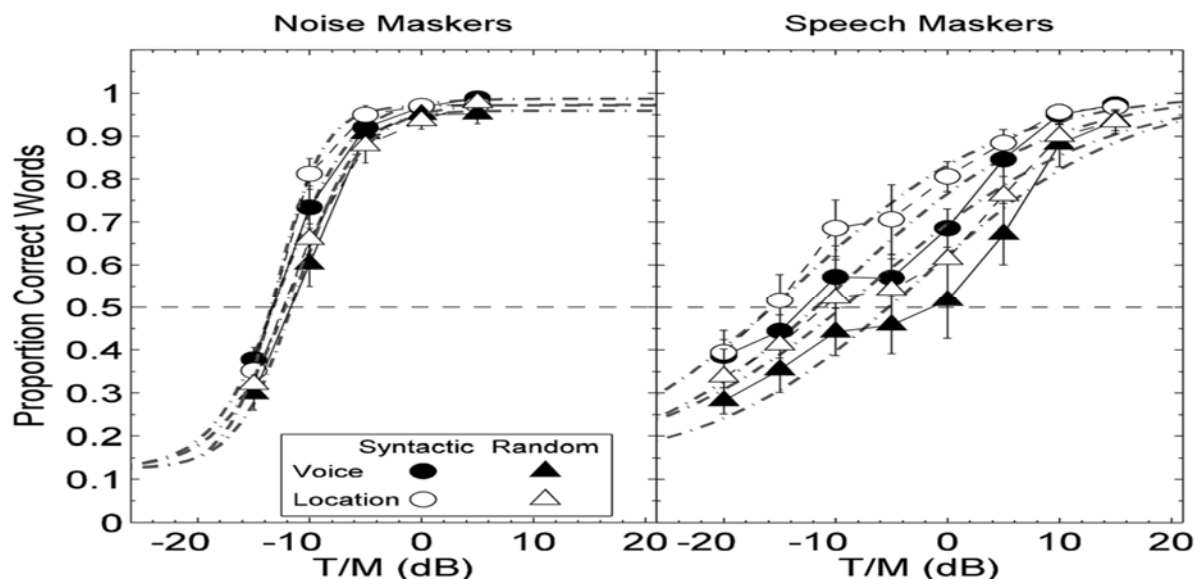
Principal findings from this study are summarized in **Figure 7** (right). These are group mean performance-level functions (proportion correct as a function of target-to-masker ratio) for a target talker and two independent noise maskers (left panel) that formed the high EM control and for two independent speech maskers that created large amounts of IM. First, note that the functions in the left panel have steep slopes and vary little across the different parameters tested (syntactically correct or not; target *voice* cue vs. target *location* cue). In fact, the differences in "thresholds" (midpoints of the functions) did not differ significantly across conditions. In contrast, the functions in the right panel for the speech maskers were much shallower consistent with a high degree of IM (Kidd et al., 1998) and the computed thresholds confirm that conformance to a correct syntax yielded significantly better performance than for incorrect target sentence syntax. This work is important because it suggests that a priori knowledge about the syntactic category of impending words in a sentence assists the listener in segregating and following over time one specific stream of speech mixed with unwanted speech sources.

III. Peer Reviewed Articles and Conference Proceedings Supported by AFOSR (2012-present)

Clayton, K.K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A.D. and Kidd, G. Jr. (2015) "Executive function, visual attention and the cocktail party problem in musicians and non-musicians," (under review)

Kidd, G. Jr., Mason, C.R., Best, V. and Swaminathan, J. (2015) "Benefits of acoustic beamforming for solving the cocktail party problem," Trends in Hearing, 19, 1-15

Swaminathan, J., Mason, C.R., Streeter, T.M., Best, V.A., Kidd, G. Jr. and Patel, A.D. (2015) "Musical training, individual differences and the cocktail party problem" Nature Scientific Reports 5, Article number: 11628 doi:10.1038/srep11628

Best, V., Mason, C.R., Kidd, G. Jr., Iyer, N. and Brungart, D.S. (2015) "Better ear glimpsing assists the listener in segregating and following efficiency in hearing-impaired listeners," J. Acoust. Soc. Am., 137, EL213-219

Kidd, G. Jr. and Mason, C.R. (2015) "Double Entendre: Embedding a secondary message in pointillistic speech," Acoust. Soc. Am. POMA, 23, 1-10

Roverud, E., Best, V., Mason, C.R. and Kidd, G. Jr. (2015) "Selective and divided listening in normal hearing and hearing-impaired listeners measured in a nonspeech pattern identification task," Acoust. Soc. Am. POMA, 23, 1-13

Best, V., Mason, C.R., Swaminathan, J., Kidd, G. Jr., Jakien, K.M., Kampel, S.D., Gallun, F.J., Buchholz, J.M. and Glyde, H. "On the contribution of target audibility to performance in spatialized speech mixtures," Proceedings of the 17th International Symposium on Hearing, Groningen, Netherlands, June, 2015, 8 pages

Kidd, G. Jr., Mason, C.R. and Best, V. (2014) "The role of syntax in maintaining the integrity of streams of speech," J. Acoust. Soc. Am., 135, 766-777

Favrot, S., Mason, C.R. and Kidd, G. Jr. (2014) "Influence Of A Microphone Array On Speech-On-Speech Masking Psychometric Functions," Proceedings of the 40th Annual German Conference on Acoustics (DAGA), Oldenburg, Germany

Kidd, G. Jr., Mason, C.R., Streeter, T., Thompson, E., Best, V. and Wakefield, G.W. (2013) "Perceiving sequential dependencies in auditory streams," J. Acoust. Soc. Am., 134, 1215-1231

Favrot, S., Mason, C.R., Streeter, T. Desloge, J. and Kidd, G. Jr. (2013) "Performance of a highly directional microphone array in a reverberant environment," Proceedings of the International. Conf. on Acoustics/Acoustical Society of America, Montreal, Canada, vol. 18, 8 pages

Best, V., Thompson, E., Mason, C.R. and Kidd, G. Jr. (2013) "Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the spectral overlap of competing talkers" J. Acoust. Soc. Am. (L), 133, 3677-3680

Swaminathan, J., Mason, C.R., Streeter, T., Best, V. and Kidd, G. Jr. (2013) "Spatial release from masking for noise-vocoded speech," Proceedings of the International Conf. on Acoustics/Acoustical Society of America, Montreal, Canada, vol. 19, 8 pages

Best, V., Thompson, E., Mason, C.R. and Kidd, G. Jr. (2013) "Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the spectral overlap of competing talkers" J. Acoust. Soc. Am. (L), 133, 3677-3680

Best, V., Thompson, E.R., Mason, C.R. and Kidd, G. Jr. (2013) "An energetic limit on spatial release from masking" J. Assoc. Res. Otolaryn., 14, 603-610

Kidd, G. Jr., Favrot, S., Desloge, J., Streeter, T., and Mason, C.R. (2013) "Design and preliminary testing of a visually-guided hearing aid," J. Acoust. Soc. Am., 133, EL202-207

Best, V., Marrone, N., Mason, C.R. and Kidd, G. Jr. (2012) "The influence of non-spatial factors on measures of spatial release from masking," J. Acoust. Soc. Am., 131, 3103-3110

IV. Papers Presented at Scientific Meetings

Clayton, K. K., Swaminathan J, Zuk J, Yazdankabahsh, A,, Patel A, and Kidd G Jr (2015) "Visual attention, executive function, and spatial hearing in musicians and non-musicians," Meeting of the Northeast Music Cognition Group, Wesleyan University, Middletown, CT

Kidd, G. Jr. and Mason, C.R. (2015) "Double entendre: embedding a secondary message in pointillistic speech," presented at the Acoustical Society of America Spring Meeting, Pittsburgh, PA

Mason, C.R., Best, V. Swaminathan, J., Roverud, E. and Kidd, G. Jr. (2015) "Application of a monaural glimpsing model to binaural speech mixtures," presented at the Acoustical Society of America Spring Meeting, Pittsburgh, PA

Best, V., Mason, C.R., Swaminathan, J., Roverud, E. and Kidd, G. Jr. (2015) "Does providing more processing time improve speech intelligibility in hearing-impaired listeners?" presented at the Acoustical Society of America Spring Meeting, Pittsburgh, PA

Roverud, E., Best, V., Mason, C.R. and Kidd, G. Jr. (2015) "Analytic and divided listening in normal-hearing and hearing-impaired listeners measured in a nonspeech pattern identification task," presented at the Acoustical Society of America Spring Meeting, Pittsburgh, PA

Kidd, G. Jr., Best, V., Desloge, J., Mason, C.R., Roverud, E., Streeter, T. and Swaminathan, J. (2015) "An audio-visual test of dynamic speech recognition using the Visually-Guided Hearing Aid (VGHA)," presented at the 38th Annual Midwinter Meeting of the Association for Reseearch in Otolaryngology, Baltimore, MD

Swaminathan, J., Mason, C.R., Streeter, T.M., Kidd, G. Jr. and Patel, A.D. (2014) "Spatial release from masking in musicians and non-musicians," J. Acoust. Soc. Am. 135 , 2281

Kidd, G. Jr., Favrot, S., Desloge, J., Streeter, T., and Mason, C.R. (2013) "Design and preliminary testing of a visually-guided hearing aid," presented at the 36th Annual Midwinter Meeting of the Association for Research in Otolaryngology, Baltimore, MD

Kidd, G. Jr., Mason, C.R. and Best, V. (2013) "The role of syntax in maintaining the integrity of streams of speech," presented at the 166th meeting of the Acoustical Society of America, San Diego, CA

Best, V., Thompson, E., Mason, C.R. and Kidd, G. Jr. (2012) "Does Energetic Masking Limit Performance in Spatialized Speech Mixtures?" Presented at the 23rd Annual Midwinter Research Meeting, Association for Research in Otolaryngology, San Diego, CA

1.

1. Report Type

Final Report

Primary Contact E-mail
Contact email if there is a problem with the report.

gkidd@bu.edu

Primary Contact Phone Number
Contact phone number if there is a problem with the report

617-353-7360

Organization / Institution name

Boston University

Grant/Contract Title
The full title of the funded effort.

Spatial hearing, attention and informational masking
in speech identification

Grant/Contract Number
AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-12-1-0171

Principal Investigator Name
The full name of the principal investigator on the grant or contract.

Gerald Dawson Kidd, Jr.

Program Manager
The AFOSR Program Manager currently assigned to the award

Patrick Bradshaw

Reporting Period Start Date

05-01-2012

Reporting Period End Date

04-30-2015

Abstract

The research completed during the past award period concerned the development and evaluation of a new means for representing the speech stimulus, called "pointillistic speech," and the role of listener expectation in speech understanding in adverse acoustic environments. The topics addressed using pointillistic processing included informational masking and the covert embedding of secondary messages in pointillistic speech and other sounds. The primary accomplishment was obtaining a better understanding of speech communication under masked conditions - especially those dominated by informational masking - while a secondary accomplishment was the testing and refinement of covert messaging using pointillistic speech. In the latter case, possible applications for increasing information transfer of audio signals and acoustic steganography were explored. These included combining natural speech with embedded pointillistic information carrying a covert secondary message that was not discernable by a human listener attending to the primary message. Under many conditions, both the primary and secondary messages were fully decipherable through listening (primary message) and signal processing (secondary message). In the area of predictability and expectation, the reslience of streams of speech to perceptual and cognitive intrusions by competing sounds was explored on a linguistic level by varying syntactic structure and by using more formal means for varying predictability by constructing target sequences of sounds from

transition matrices (i.e., Markov chains). The main finding was that listeners could extract a target stream more successfully when it contained sequences of elements that were statistically predictable as specified by the transition matrix. This approach was judged as successful in improving our understanding of the human processing of streams of speech masked by competing streams of speech.

Distribution Statement

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

Air Force SF 298 Kidd 2015.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

FINAL REPORT AFOSR 2015.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

Kidd, G. Jr., Mason, C.R., Best, V. and Swaminathan, J. (2015) "Benefits of acoustic beamforming for solving the cocktail party problem," Trends in Hearing, 19, 1-15

Swaminathan, J., Mason, C.R., Streeter, T.M., Best, V.A., Kidd, G. Jr. and Patel, A.D. (2015) "Musical training, individual differences and the cocktail party problem" Nature Scientific Reports 5, Article number: 11628 doi:10.1038/srep11628

Best, V., Mason, C.R., Kidd, G. Jr., Iyer, N. and Brungart, D.S. (2015) "Better ear glimpsing assists the listener in segregating and following efficiency in hearing-impaired listeners," J. Acoust. Soc. Am., 137, EL213-219

Kidd, G. Jr. and Mason, C.R. (2015) "Double Entendre: Embedding a secondary message in pointillistic speech," Acoust. Soc. Am. POMA, 23, 1-10

Roverud, E., Best, V., Mason, C.R. and Kidd, G. Jr. (2015) "Selective and divided listening in normal hearing and hearing-impaired listeners measured in a nonspeech pattern identification task," Acoust. Soc. Am. POMA, 23, 1-13

Best, V., Mason, C.R., Swaminathan, J., Kidd, G. Jr., Jakien, K.M., Kampel, S.D., Gallun, F.J., Buchholz, J.M. and Glyde, H. "On the contribution of target audibility to performance in spatialized speech mixtures," Proceedings of the 17th International Symposium on Hearing, Groningen, Netherlands, June, 2015, 8 pages
Kidd, G. Jr., Mason, C.R. and Best, V. (2014) "The role of syntax in maintaining the integrity of streams of speech," J. Acoust. Soc. Am., 135, 766-777

Favrot, S., Mason, C.R. and Kidd, G. Jr. (2014) "Influence Of A Microphone Array On Speech-On-Speech Masking Psychometric Functions," Proceedings of the 40th Annual German Conference on Acoustics (DAGA), Oldenburg, Germany

Kidd, G. Jr., Mason, C.R., Streeter, T., Thompson, E., Best, V. and Wakefield, G.W. (2013) "Perceiving sequential dependencies in auditory streams," J. Acoust. Soc. Am., 134, 1215-1231
Favrot, S., Mason, C.R., Streeter, T. Desloge, J. and Kidd, G. Jr. (2013) "Performance of a highly directional

microphone array in a reverberant environment," Proceedings of the International. Conf. on Acoustics/Acoustical Society of America, Montreal, Canada, vol. 18, 8 pages

Best, V., Thompson, E., Mason, C.R. and Kidd, G. Jr. (2013) "Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the spectral overlap of competing talkers" J. Acoust. Soc. Am. (L), 133, 3677-3680

Swaminathan, J., Mason, C.R., Streeter, T., Best, V. and Kidd, G. Jr. (2013) "Spatial release from masking for noise-vocoded speech," Proceedings of the International Conf. on Acoustics/Acoustical Society of America, Montreal, Canada, vol. 19, 8 pages

Best, V., Thompson, E., Mason, C.R. and Kidd, G. Jr. (2013) "Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the spectral overlap of competing talkers" J. Acoust. Soc. Am. (L), 133, 3677-3680

Best, V., Thompson, E.R., Mason, C.R. and Kidd, G. Jr. (2013) "An energetic limit on spatial release from masking" J. Assoc. Res. Otolaryn., 14, 603-610

Kidd, G. Jr., Favrot, S., Desloge, J., Streeter, T., and Mason, C.R. (2013) "Design and preliminary testing of a visually-guided hearing aid," J. Acoust. Soc. Am., 133, EL202-207

Best, V., Marrone, N., Mason, C.R. and Kidd, G. Jr. (2012) "The influence of non-spatial factors on measures of spatial release from masking," J. Acoust. Soc. Am., 131, 3103-3110

Changes in research objectives (if any):

None.

Change in AFOSR Program Manager, if any:

From Willard Larkin to Patrick Bradshaw May, 2013.

Extensions granted or milestones slipped, if any:

None.

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Oct 13, 2015 14:02:02 Success: Email Sent to: gkidd@bu.edu